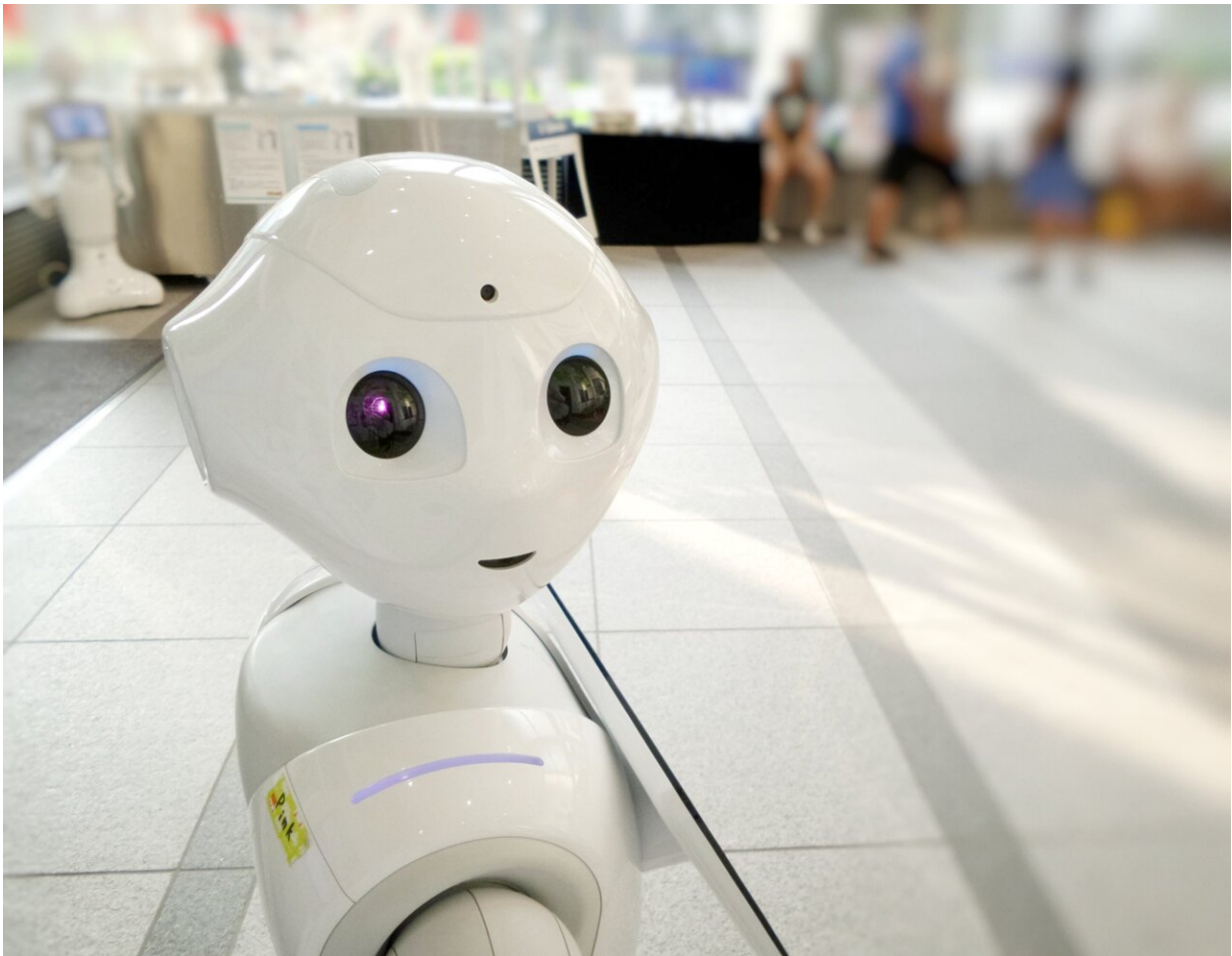


Perspective paper explores the debate over sentient machines

January 31 2024



Credit: Unsplash/CC0 Public Domain

A researcher from the New Jersey Institute of Technology has published

a perspective paper that examines sentience and its application to artificial intelligence (AI) and robotics. Sentience describes the ability to sense and feel, drawing its meaning from the Latin word *sentire* which means "to feel." The paper addresses a set of ideological commitments at stake in debates over sentient machines. The author proposes that artificial sentience is both necessary and impossible.

The perspective paper was published in the [*Journal of Social Computing*](#).

"I argue that these ideological commitments, which I call the AIdeal, create a situation where artificial sentience is both necessary and impossible. To better understand this impasse, I look at historical and contemporary discourse on artificial sentience in order to make the ideological background of the debate explicit," said Daniel Estrada, a university lecturer at New Jersey Institute of Technology.

The AIdeal is an ideological framework that implicitly structures the discourse on artificial sentience, AI, and thinking machines.

To describe this framework, Estrada reviews the historical discourse on sentience as it appears in ancient, early modern, and 20th century philosophy. He pays special attention to how these ideals are projected onto artificial agents. In AI, agents are computer programs or systems that perceive their environment and then take actions to accomplish some goal.

Estrada begins by examining sentience as an ancient tradition used to mark the distinction between plants and animals. Sentience is the capacity for sensation. So sentient creatures are sensitive to changes in their world.

Some researchers apply sensitivity to [environmental change](#) in the broadest possible sense. Their definition might include the whole of the

universe, with every molecule and particle that somehow reacts with its environment. However, most researchers use the word sentience in a more restrictive way. So the debate over sentience focuses on where to place the restrictions and draw the boundaries in the definition.

Estrada argues that tensions among these ideals result in a crisis of ideology because of the conditions where artificial sentience is both necessary and impossible. To move past this crisis does not mean that a satisfying resolution among the competing ideals has been achieved. Rather, it requires researchers to shift their focus to the material conditions and actual practices in which these ideals operate.

Estrada follows the philosopher Charles Mills in sketching a nonideal approach to AI and artificial sentience, with a looser grip on the ideology of the discourse. Estrada proposes an idea of participation that deflates the sentience discourse in AI and shifts the focus instead to the material conditions in which sociotechnical networks operate.

Artificial sentience has never been a stable goal that scientists can pursue in some objective sense, for example, in the way scientists might attempt to land on the moon or cure cancer. On the contrary, the discourse on artificial sentience looks more like an endless hallway or a carrot on a stick, where the goals are always just over the horizon.

As Estrada explains, if scientists could somehow fix the goalposts in this debate to match the ideals of earlier generations, it might reveal dozens of moon-landing scale events of technical achievement in the last fifty years. These might easily convince Aristotle, Descartes, or even Turing that an artifact had achieved sentience.

"Put simply, the prospect of artificial sentience depends less on what artifacts do, and much more on what we believe about them. That's the AIdeal," said Estrada.

Looking ahead, Estrada believes that this paper only begins to explain the ideological structure of the AI discourse. He would like to expand the discussion to cover other important areas. For example, Estrada would like to talk about the questions in the philosophy of language in the 20th century, and how this influences current debates over whether chatbots like chatGPT can "understand" language.

"Filling out this picture will contribute to ongoing debates on these important issues," said Estrada.

More information: Daniel Estrada, *AIdeal: Sentience and Ideology*, *Journal of Social Computing* (2024). [DOI: 10.23919/JSC.2023.0029](https://doi.org/10.23919/JSC.2023.0029)

Provided by Tsinghua University Press

Citation: Perspective paper explores the debate over sentient machines (2024, January 31)
retrieved 29 April 2024 from
<https://techxplore.com/news/2024-01-perspective-paper-explores-debate-sentient.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.